SDC

TM-738 001 00

SOME REMARKS ABOUT SEQUENCES IN CONTEXT FREE LANGUAGES

AFCRL-63-58

SDC Document No.

TM-738/001/00

Technical Memorandum

SDC TM SERIES

This document was produced by SDC in connection with a research project co-sponsored by SDC's independent research program and Contract AF 19(628)-485, Project No. 4641, Task No. 46410.

SCIENTIFIC REPORT NO. 2

SOME REMARKS ABOUT SEQUENCES IN CONTEXT FREE LANGUAGES

by
Seymour Ginsburg
System Development Corporation
Joseph S. Ullian
University of Chicago

31 January 1963

Prepared for
ELECTRONICS RESEARCH DIRECTORATE
AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS

SYSTEM DEVELOPMENT CORPORATION
2500 COLORADO AVE.
SANTA MONICA
CALIFORNIA

Requests for additional copies by Agencies of the Department of Defense, their contractors, and other government agencies should be directed to:

ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

Department of Defense contractors must be established for ASTIA services or have their "need-to-know" certified by the cognizant military agency of their project or contract.

All other persons and organizations should apply to:

U. S. DEPARTMENT OF COMMERCE
OFFICE OF TECHNICAL SERVICES
WASHINGTON 25, D. C.

ABSTRACT

The following conjecture is considered:

(*) It is unsolvable whether a language (= context free language) contains a sequence.

While this conjecture is left unresolved, a number of results pertaining to it are obtained. For example, the unsolvability of whether a language contains the set $\{aba^2ba^3\ldots ba^n/n \geq 1\}$, implies (*). It is shown that (*) is equivalent to the unsolvability of whether a language contains a chain of a special form. Several facts about whether a language contains a specific sequence are also demonstrated. In particular, it is shown that whether a language contains a given sequence is unsolvable, but whether a language contains a given ultimately periodic sequence is solvable.

SOME REMARKS ABOUT SEQUENCES IN CONTEXT FREE LANGUAGES

Introduction

In [6] it was shown by a complicated argument that for two (context free) languages L_1 and L_2, it is recursively unsolvable whether there exists a complete sequential machine mapping L_1 into L_2. Now an alternative (and quite simple) proof of this fact would follow from verification of the following conjecture: It is recursively unsolvable whether a language contains an ultimately periodic sequence. [For the language $\{a^n/n \geq 1\}$ can be mapped into an arbitrary language L by a complete sequential machine if and only if L contains an ultimately periodic sequence.] Neither this conjecture nor its analogue for sequences in general has been settled, but they have provided motivation for the study of sequences in languages. The present paper sets forth several results about sequences in languages and shows how some of the questions which remain unanswered may be reduced.

The paper is divided into four sections. Section 1 reviews the terminology of languages. In section 2 a language is exhibited which contains exactly one sequence, a sequence which is not ultimately periodic. Using this language it is shown that the unsolvability of whether a language contains a set of words of a certain form implies the unsolvability of whether a language contains a sequence. In section 3 the solvability of a language containing a sequence is proved equivalent to the solvability of a language containing at least one special kind of chain. In section 4 it is shown that whether a language contains a given sequence is unsolvable, but whether a language contains a given ultimately periodic sequence is solvable.

Section 1. Preliminaries

Let Σ be a finite nonempty set and let $\theta(\Sigma)$ be the free semigroup with identity ϵ generated by Σ. (Thus $\theta(\Sigma)$ is the set of all finite sequences, or words, of Σ and ϵ is the empty sequence.) We shall be considering certain subsets of $\theta(\Sigma)$ which are called "context free languages," or "languages" for short. These languages arose in the study of natural languages [2] and have been shown to be identical with the components in the "ALGOL-like" artificial languages which occur in data processing [4].

A grammar G is a 4-tuple (V,P,Σ,S), where V is a finite set, Σ is a subset of V, S is an element of $V-\Sigma$, and P is a finite set of ordered pairs of the form (ξ,w) with ξ in $V-\Sigma$ and w in $\theta(V)$. P is called the set of productions of G. An element (ξ,w) in P is denoted by $\xi \rightarrow w$. If x and y are in $\theta(V)$, then we write $x \Rightarrow y$ if either $x = y$ or there exists a sequence $x = x_1, x_2,\ldots,x_n = y$ $(n > 1)$ of elements in $\theta(V)$ with the following property. For each $i < n$ there exist a_i, b_i, ξ_i, w_i such that $x_i = a_i\xi_i b_i$, $x_{i+1} = a_i w_i b_i$, and $\xi_i \rightarrow w_i$. The language generated by G, denoted by L(G), is the set of words $\{w/S \Rightarrow w, w \text{ in } \theta(\Sigma)\}$. A context free language (over Σ) is a language L(G) generated by some grammar $G = (V,P,\Sigma,S)$. Unless otherwise stated, by a language we shall always mean a context free language.

If A and B are subsets of $\theta(\Sigma)$, then the set of words $\{ab/a \text{ in } A, b \text{ in } B\}$ is called the product of A and B and is written AB. If A (or B) consists of just one word, say $A = \{a\}$ $(B = \{b\})$, then aB (Ab) is written instead of AB.

For each word x in $\theta(\Sigma)$, $|x|$ denotes the length of x.

If A and B are languages, then so are AB, $A + B$,[1] and A^*[2] [3].

The family of regular sets is characterized as the smallest family of subsets of $\theta(\Sigma)$ containing the finite sets and closed under the operations of union, product, and * [9]. Each regular set is a language [3].

Let $x_1, \ldots, x_n, \ldots$ (written $x_1 \ldots x_n \ldots$) be an infinite sequence of elements of Σ. The sequence is said to be contained in a set H of words (or H contains the sequence) if the word $x_1 \ldots x_i$ is in H for each i. A sequence $x_1 \ldots x_n \ldots$ is said to be ultimately periodic (u.p.) if there exist integers n_0 and p so that $x_{n+p} = x_n$ for $n \geq n_0$. An infinite sequence of words $\{w_i\}$ is said to be a chain if w_i is an initial subword of w_{i+1} for each i, i.e., $w_{i+1} = w_i u_i$ for some u_i in $\theta(\Sigma)-\epsilon$. A language is said to contain a chain if the language contains each word in the chain.

We are interested in whether a language containing an u.p. sequence is unsolvable and whether a language containing a sequence is unsolvable; and we shall present a number of results which have arisen during a study of these two problems.

---

(1) Both "+" and "U" are used to denote set union.

(2) If A is a set of words, then $A^* = \epsilon + \bigcup_{1}^{\infty} A^n$, where $A^1 = A$ and $A^{i+1} = A^i A$.

Section 2. Distinguished Sequences

Consider the question of whether or not a language containing a sequence must contain an u.p. sequence. By application of a systematic procedure, we can effectively enumerate those languages which contain no sequences. Since we can test a given language to see if it contains a specified u.p. sequence,[3] and since we can effectively enumerate the u.p. sequences, we also have a systematic procedure for effectively enumerating those languages which contain u.p. sequences. Therefore, if each language containing a sequence contained an u.p. sequence, we would have a decision procedure for determining whether or not a given language contained an (u.p.) sequence. However, we now show that there are languages which contain a sequence but no u.p. sequence.

Notation. Given a word w and an element b in Σ, let $\#b(w)$ be the number of occurrences of b in w.

Theorem 2.1. There exists a language which contains a sequence but no u.p. sequence.

Proof. Let $G_1 = (V_1, P_1, \Sigma, \xi)$, where $\Sigma = \{a,b\}$, $V_1 = \xi \cup \Sigma$, and $P_1 = \{\xi \to b, \xi \to a\xi, \xi \to b\xi a\}$. Let $L_1 = L(G_1)$. Clearly $L_1 = \{u/u = wba^{\#b(w)}, w \text{ in } \theta(a,b)\}$. If $u = wba^n$ is in L_1, then wba^nba^{n+1} is in L_1 and is the proper extension of u in L_1 of smallest length. Hence L_1 contains the chain C = b, bba, $bbaba^2$, $bbaba^2ba^3, \ldots$.

---

[3] See Theorem 4.2.

Let $G_2 = (V_2, P_2, \Sigma, \xi)$, where $V_2 - \Sigma = \{\xi, \nu, \gamma\}$ and $P_2 = \{\xi \to \nu\gamma,\ \gamma \to ba,$ $\gamma \to a\gamma,\ \gamma \to b\gamma,\ \gamma \to b\gamma a,\ \nu \to b,\ \nu \to a\nu,\ \nu \to b\nu,\ \nu \to \nu a\}$. Let $L_2 = L(G_2)$. Then

$$L_2 = \{u/u = wba^n,\ 1 \le n \le \#b(w),\ w \text{ in } \theta(a,b)\, b\theta(a,b)\}.$$

Let $L_3 = b + L_1 b + L_2$. Note that each word in $b + L_1 b$ ends in b, and each word in L_2 ends in a. Let D be the sequence $bbaba^2ba^3b\ldots$. Obviously D is not u.p. We shall show that L_3 contains D but no other sequence.

Each word in D ending in b is in $b + L_1 b$. Since each word in D ending in a is in L_2, D is contained in L_3. Now let E be any sequence contained in L_3. Neither a nor ba is in L_2, thus neither is in L_3. Therefore E begins with bb. Now suppose that E begins with $buba^n$, $n \ge 0$, for some word u in $\theta(a,b)$. Two cases arise.

(α) $n = \#b(bu)$. Then $buba^{n+1}$ is not in L_2, and thus not in L_3. Hence E must begin with $buba^n b$.

(β) $n < \#b(bu)$. Since $n \neq \#b(bu)$, $buba^n$ is not in L_1. Thus E cannot begin with $buba^n b$, that is, E begins with $buba^{n+1}$.

By induction it therefore follows that E = D. Hence D is the only sequence contained in L_3.

Using the languages constructed in Theorem 2.1 we now show that if either of two problems is recursively unsolvable, then so is the problem of whether a given language contains a sequence.

Let D, L_1, L_2, and L_3 be as in Theorem 2.1 and let $\theta = \theta(a,b)$. For M and N subsets of θ, let $\tau(M,N) = b + L_1 b + \bigcup_{n=0}^{\infty} (Nb)^{n+1} Mba^{n+1}$. It is readily seen that $\tau(M,N)$ is a language if M and N are. Furthermore, $\tau(M_1,N_1) \subseteq \tau(M,N)$ if $M_1 \subseteq M$ and $N_1 \subseteq N$. Then

$$\begin{aligned}\tau(\theta,\theta) &= b + L_1 b + \bigcup_{n=0}^{\infty} (\theta b)^{n+2} a^{n+1} \\ &= b + L_1 b + L_2 \\ &= L_3.\end{aligned}$$

Since D is the only sequence in L_3, it follows that $\tau(M,N)$ contains a sequence if and only if it contains the sequence D.

Consider $\tau(M,a^*) = b + L_1 b + \bigcup_0^{\infty} (a^* b)^{n+1} Mba^{n+1}$. If this set is to contain D then it is necessary that M (i) contain each of the words ϵ, ba, $baba^2$, $baba^2ba^3, \ldots$ (to obtain members in the sequence of the form wba); and (ii) contain each word of the form $a^{n+1}ba^{n+2}\ldots ba^{n+m}$, $n \geq 0$, $m \geq 1$ (to obtain members in the sequence of the form wba^{n+2}, with $m = \#b(w)-n-1$). Moreover, if M satisfies (i) and (ii), then it is readily seen that $D \subseteq \tau(M,a^*)$ since $D \cap \theta b \subseteq b + L_1 b$.

Suppose that $M = bP + a\theta + \epsilon$. Then (ii) is satisfied; and (i) is satisfied if and only if P contains the set $\{aba^2\ldots ba^n / n \geq 1\}$. Thus $\tau(M,a^*)$ contains a sequence if and only if P contains $\{aba^2\ldots ba^n / n \geq 1\}$. Therefore we get

<u>Theorem 2.2.</u> If whether a language contains $\{aba^2\ldots ba^n / n \geq 1\}$ is unsolvable, then whether a language contains a sequence is unsolvable.

Suppose that $M = P + a^{*} + b\theta$. Then (i) is satisfied, and (ii) is satisfied for $m = 1$. Now (ii) holds for every i if and only if P contains each word of the form $a^{i}ba^{i+1}\ldots ba^{i+k}$, $i, k \geq 1$. Thus we have

<u>Theorem 2.3</u>. If whether a language contains $\{a^{i}ba^{i+1}\ldots ba^{i+k}/i, k \geq 1\}$ is unsolvable, then whether a language contains a sequence is unsolvable.

We now consider sequences D with the property that there is a language containing D and no other sequence.

<u>Definition</u>. A sequence D of words with the above property is called a <u>distinguished</u> sequence.

Since every u.p. sequence is a language, each u.p. sequence is distinguished. The sequence D in Theorem 2.1 shows that the converse is not true, i.e., there are distinguished sequences which are not u.p.

Given a distinguished sequence D we may obtain other distinguished sequences as follows. Let S be any complete sequential machine[4] with the

---

[4]A <u>generalized</u> <u>sequential</u> <u>machine</u> S is a 6-tuple $(K,\Sigma,\Delta,\delta,\lambda,p_1)$ where (i) K is a finite nonempty set (of "states"); (ii) Σ is a finite nonempty set (of "inputs"); (iii) Δ is a finite nonempty set (of "outputs"); (iv) δ is a mapping of $K \times \Sigma$ into K (the "next state" function); (v) λ is a mapping of $K \times \Sigma$ into $\theta(\Delta)$ (the "output" function); and (vi) p_1 is an element of K (the "start" state). A <u>complete</u> <u>sequential</u> <u>machine</u> is a generalized sequential machine in which λ maps $K \times \Sigma$ into Δ.

property that at each state, λ maps Σ one to one into Δ. Let L be a language containing a distinguished sequence D. Then $S(L)$[5] is a language [5] containing the sequence S(D). That S(L) contains no sequence but S(D) follows from λ mapping Σ one to one into Δ. Furthermore, if D is not u.p., neither is S(D). We omit the straightforward details.

The question naturally arises: Are there any sequences which are not distinguished? A simple cardinality argument shows there are. For there are $2^{\aleph_0}$ sequences when Σ contains at least two elements, and only $\aleph_0$ languages. Thus there exists a sequence D (in fact $2^{\aleph_0}$) such that any language containing D contains at least one other sequence, i.e., a sequence D which is not distinguished. More precisely,

<u>Theorem 2.4.</u> Every distinguished sequence is recursive.

<u>Proof</u>. Let $L \subseteq \theta(a_1,\ldots,a_k)$ be a given language. We outline an effective procedure P with the following property: If L contains at least one sequence $\{w_i\}_{i \geq 1}$ and every sequence in L contains w_n, then P selects w_n at the n-th stage.

Each stage of P is divided into substages. In substage $m(\geq 1)$ of stage 1, P determines the finite set

---

(5) Extend δ and λ to $K \times \theta(\Sigma)$ as follows. Let $\delta(q,\epsilon) = q$ and $\lambda(q,\epsilon) = \epsilon$. For each word $x_1\ldots x_{k+1}$, each x_i in Σ, let $\delta(q,x_1\ldots x_{k+1}) = \delta[\delta(q,x_1\ldots x_k), x_{k+1}]$ and $\lambda(q,x_1\ldots x_{k+1}) = \lambda(q,x_1\ldots x_k)\,\lambda[\delta(q,x_1\ldots x_k), x_{k+1}]$. For each word w, let $S(w) = \lambda(p_1,w)$. For each set L, let $S(L) = \{S(w)/w \text{ in } L\}$.

$$D_{1,m} = \{w/\ |w| = m \text{ and Init } w^{(6)} \subseteq L\}.$$

If $D_{1,m}$ is empty, P selects a_k and proceeds to stage 2. If $D_{1,m}$ is nonempty and each word in it begins with the same letter, say b_1, P selects b_1 and proceeds to stage 2. Otherwise P proceeds to substage m + 1.

If L contains at least one sequence and every sequence in L contains b_1, then there must be some $m \geq 1$ such that $D_{1,m}$ is nonempty and contains only words beginning with b_1. (This follows immediately from the well-known Infinity Lemma of graph theory [7;p.81]). In this case P completes its first stage and selects b_1.

Suppose that P completes the n^{th} stage $(n \geq 1)$ singling out $b_1 \dots b_n$. In substage m (≥ 1) of stage n + 1, P determines the finite set

$$D_{n+1,m} = \{w/\ |w| = n + m, \text{ Init } w \subseteq L, \text{ and } b_1 \dots b_n \text{ is in Init } w\}.$$

If $D_{n+1,m}$ is empty, P selects $b_1 \dots b_n a_k$ and proceeds to stage n + 2. If $D_{n+1,m}$ is nonempty and every word in it has the same $n + 1^{st}$ letter, say b_{n+1}, then P selects $b_1 \dots b_n b_{n+1}$ and proceeds to stage n + 2. Otherwise P proceeds to substage m + 1. If L contains at least one sequence and every sequence in L contains $b_1 \dots b_{n+1}$, then there must be some $m \geq 1$ such that $D_{n+1,m}$ is nonempty and contains only words with b_{n+1} as its $n + 1^{st}$ letter. In this case P completes its $n + 1^{st}$ stage and selects $b_1 \dots b_{n+1}$.

---

(6) For a word w, Init $w = \{u/u \neq \epsilon,\ uv = w \text{ for some } v\}$. For a set H of words let Init $H = \bigcup_{w \text{ in } H} \text{Init } w$.

Suppose that L contains exactly one sequence D. Then P completes the n^{th} stage for every n and enumerates D. Since P is an effective procedure, D must be recursively enumerable. But a sequence is recursively enumerable if and only if it is recursive. Hence the result.

The next theorem shows the existence of recursive sequences which are not distinguished. Since the proof involves special concepts, it is given in the appendix.

Theorem 2.5. Let a be a given element of Σ. Then each recursive, non u.p. sequence D with the property that for every $n \geq 1$ there is a word ua^k, $u \neq \epsilon$, in D such that $k \geq 2^{n|u|}$ is not distinguished.[7]

In passing, we mention two open problems.

(1) Characterize the set of distinguished sequences.

(2) Characterize the set of those sequences D having the property that there exists a language containing D but no u.p. sequence.

Section 3. An Equivalence Condition

We now show that the solvability of a language containing a sequence is equivalent to the solvability of a language containing a special kind of chain.

---

[7] One such sequence $D = x_1 \ldots x_n \ldots$ is obtained by letting $f(0) = 1$, $f(n+1) = f(n) + 2^{(n+1)f(n)} + 1$ for $n \geq 0$, $x_i = b$ if i is in the range of f, and $x_i = a$ otherwise.

<u>Theorem 3.1</u>. Call C(P) a counting chain, where P is a subset of the positive integers, if C(P) is the set of those words

$$b^{i_1}a^{i_1}b^{i_2}a^{i_1+i_2}\ldots b^{i_k}a^{\sum_{j=1}^{k} i_j},$$

$k > 1$, such that for $1 \le j \le k$, $i_j = 2$ if j is in P and $i_j = 1$ if j is not in P. Then the question of whether an arbitrary language contains a sequence is solvable if and only if the question of whether an arbitrary language contains a counting chain is solvable.

<u>Proof</u>. (1) Let σ be the operation which takes each occurrence of b into $\{b,b^2\}$, and leaves a unchanged. That is, $\sigma(\epsilon) = \epsilon$ and $\sigma(x_1\ldots x_r) = \sigma(x_1)\ldots\sigma(x_r)$, where $\sigma(a) = a$ and $\sigma(b) = \{b,b^2\}$. Let τ be the operation which takes each occurrence of a into $A_1 = \{ba^{2n}/n \ge 1\}$ and each occurrence of b into $A_2 = \{ba^{2n+1}/n \ge 0\}$. Since A_1 and A_2 are languages, σ and τ preserve languages, i.e., if L is a language, so are $\sigma(L)$ and $\tau(L)$ [1]. We shall show that for a subset L of θ, $\sigma\tau(L)$ contains a counting chain if and only if L contains a sequence. Therefore if it is solvable whether an arbitrary language contains a counting chain, then it is solvable whether an arbitrary language contains a sequence.

To this end let L be a subset of θ. Suppose that L contains a sequence $D = x_1\ldots x_n\ldots$. For each n let $d_n = x_1\ldots x_n$. Let $A_3 = \{1\}$ if $x_1 = a$ and $A_3 = \varphi$ if $x_1 = b$. Let

$$P = A_3 \cup \{n/n \ge 2,\ x_{n-1} = x_n\}.$$

For each n, $ba^{i_1}ba^{i_2}b\ldots ba^{i_n}$ is in $\tau(d_n)$ if and only if

$$\{j/j \le n,\ i_j > 0 \text{ and even}\} = \{j/j \le n,\ x_j = a\}.$$

For each n let u_n be the element in $\tau(d_n)$ with $i_1 = 1$ or 2 and $i_{j+1} = i_j + 1$ or $i_{j+1} = i_j + 2$, $1 \le j < n$. Then

$$\sigma(u_n) = \{b^{k_1}a^{i_1}b^{k_2}\ldots b^{k_n}a^{i_n}/k_j = 1,2; 1 \le j \le n\}.$$

In particular, $\sigma(u_n)$ contains an element v_n in which $k_j = 2$, $1 \le j \le n$, if and only if j is in P. Thus $k_1 = 2$ if and only if $x_1 = a$. Hence $k_1 = i_1$. Furthermore, for each j, $k_{j+1} = 2$ if and only if $x_j = x_{j+1}$, whence $k_{j+1} = i_{j+1} - i_j$. Thus $\{v_i\}_{i \ge 1}$ is the counting chain C(P). Since $\{v_i\} \subseteq \sigma(\{u_n\}_{n \ge 1})$, $\{u_n\}_{n \ge 1} \subseteq \tau(D)$, and $D \subseteq L$, it follows that $C(P) \subseteq \sigma\tau(L)$.

Now suppose that the counting chain $C(P) \subseteq \sigma\tau(L)$ for some set P of positive integers. For each n let

$$v_n = b^{i_1}a^{i_1}\ldots b^{i_n}a^{\sum_{j=1}^{n} a_i}$$

be in C(P). Then there is a unique word d_n in L such that v_n is in $\sigma\tau(d_n)$. In fact, d_n is the word $x_1 \ldots x_n$ of length n in $\theta(a,b)$ for which (i) $x_1 = a$ if and only if $i_1 = 2$, and (ii) for $1 \le j < n$, $x_j = x_{j+1}$ if and only if $i_{j+1} = 2$. It readily follows that $\{d_i\}_{i \ge 1}$ is a sequence in L.

(2) We shall now show that if it is solvable whether an arbitrary language contains a sequence, then it is solvable whether an arbitrary language contains a counting chain. This and (1) will then imply Theorem 3.1.

Let L_1 be the same as in Theorem 2.1. Let C(P) be an arbitrary counting chain. If u is a word in C(P), then from the definition of counting chain,

$u = vba^{\#b(vb)}$ for some word v in $\theta(a,b)$. Thus bu is in L_1 for each word u in $C(P)$. Hence $bC(P) \subseteq L_1$.

Let M be a given language over $\Sigma = \{a,b\}$. Let

$$H_1 = \{b(\theta b)^n a^*(b+b^2+b^2a)a^{n+1}/n \geq 1\},$$

$$H_2 = ba^*(b+b^2+b^2a)a + bMba + bMb^2a,$$

and $\bar{M} = H_1 + H_2.$

As is easily seen, H_1, H_2, and $\bar{M}$ are languages. H_1 consists of all words of the form

(α) $buba^m b^t a^p$, where $t = 1$ or 2, u is in θ, $m \geq 0$, and $2 \leq p < \#b(buba^m b^t)$.

H_2 consists of all words of the form

(β) $ba^m b^t a$ or $ba^m b^2 a^2$ for $t = 1, 2$, and $m \geq 0$;

and (γ) $bMb^t a$ for $t = 1, 2$.

Consider the language $M_1 = b + b^2 + b^3 + L_1(b+b^2) + \bar{M}$. We shall show that M_1 contains a sequence if and only if M contains a counting chain. Therefore if it is solvable whether an arbitrary language contains a sequence, it is solvable whether an arbitrary language contains a counting chain.

Suppose that vb is an initial subword of a counting chain $C(P)$. Then either (i) $v = \epsilon$, (ii) $v = b$, (iii) v is in $C(P)$, or (iv) $v = v_1 b$ with v_1 in $C(P)$. Consider bvb. If (i) holds, then $bvb = b^2$. If (ii) holds, then $bvb = b^3$. If (iii) holds, then bv is in L_1, so that bvb is in $L_1 b$. If (iv) holds, then bv_1 is in L_1, so that $bvb = bv_1b^2$ is in L_1b^2. Thus if vb is an initial subword of a counting chain, then $b^2 + b^3 + L_1(b+b^2)$ contributes bvb to M_1.

Consider initial subwords va of a counting chain. It is clear that only $\bar{M}$ can contribute words ending in a to M_1. Suppose that va is an initial subword of a counting chain. Then $bva = bwb^ta^q$, where $t = 1$ or 2, w is a word not ending in b, and $0 < q < \#b(bwb^t)$. If w contains b and $q \geq 2$, then bva is in $\bar{M}$ by (α). If w does not contain b and either $q = 1$ or $q = t = 2$, then bva is in $\bar{M}$ by (β). Thus neither (α) nor (β) contribute bva to $\bar{M}$ if and only if $bva = bw_1bw_2b^ta$ for some words w_1 and w_2, w_2 not ending in b, $t = 1$ or 2.

Suppose that M contains a counting chain C(P). By the previous discussion, M_1 contains bz for every initial subword z of C(P) except possibly $z = w_1bw_2b^ta$, $t = 1$ or 2, w_2 a word ending in a. In this case, w_1bw_2 is in C(P), thus in M. Then by (γ), bz is in $\bar{M}$, thus in M. Therefore M_1 contains bz for every initial subword z of C(P). Since b is also in M_1, M_1 contains every initial subword of bC(P). Thus M_1 contains a sequence.

Now suppose that M_1 contains a sequence D. It is easily seen that D begins with b^2. Since neither b^2 nor b^3 is in L_1, b^4 is not in M_1. Since b^2a^2 is not in $\bar{M}$, b^2a^2 is not in M_1. Thus D begins with either b^2ab or b^3a.

(a) Suppose that D begins with b^2ub^2, where u is a word ending in a. Since each word in L_1 ends in a, neither b^2ub nor b^2ub^2 is in L_1. Thus b^2ub^3 is not in M_1, so that D begins with b^2ub^2a.

(b) Suppose that D begins with $b^2ub^ta^q$, where $t = 1$ or 2, $q > 0$, and u is a word not ending in b (thus u might be ϵ). Two alternatives arise.

(i) $q = \#b(bub^t)$. Suppose that D begins with $b^2ub^ta^{q+1}$. Then $q + 1 \geq 3$, so that $b^2ub^ta^{q+1}$ can only be contributed to M_1 by (α). Since $q + 1 = \#b(b^2ub^t)$, $b^2ub^ta^{q+1}$ cannot be contributed by (α). Thus D begins with $b^2ub^ta^qb$.

(ii) $q \neq \#b(bub^t)$. Since D is a sequence, it follows from (i) that $q < \#b(bub^t)$. Suppose that D begins with $b^2ub^ta^qb$. Then $b^2ub^ta^q$ is in L_1, so that $q = \#b(bub^t)$, a contradiction. Therefore D begins with $b^2ub^ta^{q+1}$.

From (a) and (b) and the fact that D begins with b^2ab or b^3a, it follows that D consists of all initial subwords of $bC(P)$ for some counting chain $C(P)$, with $i_1, i_2, \ldots$ the exponents of b's in the chain.

Finally, suppose that M does not contain $C(P)$. In particular, let

$$x = b^{i_1}a^{i_1}\ldots b^{i_k}a^{\sum_1^k i_j}$$

be in $C(P)$ but not M. Consider $y = bxb^{i_{k+1}}a$. Clearly y is not contributed to M_1 by either (α) or (β). Since x is not in M, y is not contributed to M_1 by (γ). Thus y is not in M_1, contradicting the assumption that M_1 contains D. Thus M contains $C(P)$, i.e., M contains a counting chain if and only if M_1 contains a sequence. Q.E.D.

<u>Remark.</u> While the problem of whether an arbitrary language contains a counting chain is, as yet, unresolved, the problem of whether an arbitrary language contains <u>all</u> counting chains is recursively unsolvable. For let $\Sigma = \{a,b\}$ and let

σ and τ be as in Theorem 3.1. For each language L, it is readily seen that $\sigma\tau(L)$ contains all counting chains if and only if $\theta(a,b) - \epsilon \subseteq L$. Since $L = \theta(a,b)$ is recursively unsolvable [1], it follows that whether $\theta(a,b) - \epsilon \subseteq L$ is recursively unsolvable. Thus whether $\sigma\tau(L)$, hence an arbitrary language, contians all counting chains is recursively unsolvable.

Section 4. Special Sequences

In this section we consider whether a language L contains a specific sequence is solvable. We exhibit one set of sequences for which it is unsolvable and another for which it is solvable. We then show that whether a language contains a sequence of a special form is unsolvable.

Theorem 4.1. Given a sequence D and language L, it is recursively unsolvable whether $D \subseteq L$.

Proof. Let F be the set of all sequences of the form

$$\text{Init}\ (cw_1cw_2cw_3\ldots),$$

where $\bigcup_{i=1}^{\infty} w_i = \theta(a,b)$. Given a language M, consider $L(M) = \text{Init}\ [(cM)^*]$.(8) If $M = \theta(a,b)$, then $D \subseteq L(M)$ for every D in F. If $M \neq \theta(a,b)$, then $D \subseteq L$ for no D in F. Since it is recursively unsolvable whether $M = \theta(a,b)$ for an arbitrary language M [1], it is recursively unsolvable whether $D \subseteq L$ for some specific (or even one) D in F.

---

(8) If L is a language, then Init L is a language [5].

We next show that it is solvable whether a language contains a given u.p. sequence.

<u>Lemma 4.1</u>. Given words w_1, w_2, w_3, it is recursively solvable whether an arbitrary language L contains $w_1w_2^*w_3$.

<u>Proof</u>. If $w_2 = \epsilon$, then it is recursively solvable whether w_1w_3 is in L. Suppose that $w_2 \neq \epsilon$. Since $w_1w_2^*w_3$ is regular and L is a language, $A = w_1w_2^*w_3 \cap L$ is a language and effectively calculable from L [1]. Since $w_1w_2^*w_3 \subseteq L$ if and only if $A = w_1w_2^*w_3$, it suffices to show that whether A and $w_1w_2^*w_3$ are equal is solvable. Let τ_1 (τ_2) be the operation which maps a word x into x_1 if $x = w_1x_1$ $(x = x_1w_3)$ and into φ otherwise. Then A and $w_1w_2^*w_3$ are equal if and only if $\tau_2\tau_1(A) = w_2^*$. Now $\tau_2\tau_1(A)$ is a language and effectively calculable from A [5]. Since $w_2 \neq \epsilon$, $w_2 = y_1 \cdots y_r$, y_i in Σ. Consider the generalized sequential machine $S = (K, \Sigma, \{a\}, \delta, \lambda, p_1)$, where $K = \{p_1, \ldots, p_r\}$, $\lambda(p_i, y) = \epsilon$ for $i \neq r$, $\lambda(p_r, y) = a$, $\delta(p_i, y) = p_{i+1}$ for $i < r$, and $\delta(p_r, y) = p_1$, y in Σ. Then $S[\tau_2\tau_1(A)] = \{a^k / w_2^k \text{ in } \tau_2\tau_1(A)\}$. Since $\tau_2\tau_1(A)$ is a language, $S[\tau_2\tau_1(A)]$ is a language and effectively calculable from $\tau_2\tau_1(A)$ and S [5]. From [4], a language on one letter is a regular set and is effectively calculable as a regular set. But $A = w_1w_2^*w_3$ if and only if $S[\tau_2\tau_1(A)] = a^*$. Now it is solvable whether two regular sets are equal [9]. Thus it is solvable whether $S[\tau_2\tau_1(A)] = a^*$. Hence the result.

<u>Theorem 4.2</u>. Given an u.p. sequence D and a language L, it is solvable whether L contains D.

Proof. Let D be an u.p. sequence. Then $D = \text{Init}\ (w(a_1 \ldots a_p)^*)$, a_i in Σ, for some word w and some $p \geq 1$. For any language L, $D \subseteq L$ if and only if L contains each of the following p + 1 sets: Init w, $w(a_1 \ldots a_p)^*$, $wa_1(a_2 \ldots a_p a_1)^*$, $wa_1a_2(a_3 \ldots a_p a_1 a_2)^*, \ldots, wa_1 \ldots a_{p-1}(a_p a_1 \ldots a_{p-1})^*$. Inclusion of Init w is solvable because Init w is finite and a language is a recursive set. Each of the other inclusions is solvable by Lemma 4.1. Thus whether L contains D is solvable.

We have just seen that whether a language contains a specific u.p. sequence is solvable. The question arises as to whether a language contains at least one u.p. sequence with a given period is solvable. We now show that it is not.

Theorem 4.3. Given a language L and non-ϵ word w, it is recursively unsolvable whether L contains an u.p. sequence with period w, that is, a sequence of the form Init (w_1w^*).

Proof. Let a, b, and c be three letters not occurring in w. For each positive integer j denote by $\overline{j}$ the word ab^j. For each n-tuple $z = (z_1, \ldots, z_n)$ of words in $\theta(a,b)-\epsilon$ let

$$L(z) = \{z_{i_k} \ldots z_{i_1} c\ \overline{i_1} \ldots \overline{i_k} / k \geq 1, \text{ each } i_j \leq n\}.$$

Let $A(w) = \epsilon + \text{Init}\ w - w$. For any two n-tuples $x = (x_1, \ldots x_n)$ and $y = (y_1, \ldots, y_n)$ of words in $\theta(a,b)-\epsilon$ let

$$L(x,y) = \text{Init}\ L(x) + L(x)\ w(w^2)^* A(w) + L(y)\ (w^2)^* A(w).$$

Suppose that $L(x) \cap L(y) \neq \varphi$. Let w_1 be an element in $L(x) \cap L(y)$. Then Init $(w_1w^*) \subseteq L(x,y)$, so that L(x,y) contains an u.p. sequence with period w.

Suppose that $L(x,y)$ contains an u.p. sequence with period w. Then there is a word w_1 such that Init $(w_1w^*) \subseteq L(x,y)$. In particular, w_1w and w_1w^2 are in $L(x,y)$. Since $L(x) \subseteq \theta(a,b,c)$ and w is not in $\theta(a,b,c)$, neither w_1w nor w_1w^2 is in Init $L(x)$. Suppose that w_1w and w_1w^2 both are in $L(x)w(w^2)^*A(w)$. Since a, b, and c do not occur in w, $w_1w = uvw$ and $w_1w^2 = uvw^2$, with u in $L(x)$ and vw, vw^2 in $w(w^2)^*A(w)$. Now for each word t in $w(w^2)^*A(w)$

$$(*) \quad |t| = (2r_t + 1)\ |w| + s_t,$$

where r_t is a nonnegative integer, and $0 \le s_t \le |w| - 1$. Thus $|vw| = (2r_{vw} + 1)$ $|w| + s_{vw}$. Then

$$\begin{aligned} |vw^2| &= |vw| + |w| \\ &= (2r_{vw} + 1)\ |w| + s_{vw} + |w|. \end{aligned}$$

Since $0 \le s_{vw} \le |w| - 1$, $|vw^2|$ is not of the form given in (*). Therefore one of the words w_1w or w_1w^2 is not in $L(x)w(w^2)^*A(w)$. Similarly one of the words w_1w or w_1w^2 is not in $L(y)(w^2)^*A(w)$. Thus w_1w is in one of the two sets, $L(x)w(w^2)^*A(w)$ or $L(y)(w^2)^*A(w)$, and w_1w^2 is in the other. Therefore $w_1 = u_1v_1 = u_2v_2$, with u_1 in $L(x)$, u_2 in $L(y)$, and v_1, v_2 in Init $(w^*) + \epsilon$. Since u_1u_2 and v_1v_2 have no common letters, $u_1 = u_2$ and u_1 is in $L(x) \cap L(y)$. Thus $L(x) \cap L(y) \ne \varphi$.

It thus follows that a necessary and sufficient condition that $L(x,y)$ contain an u.p. sequence with period w is that $L(x) \cap L(y) \ne \varphi$. But $L(x) \cap L(y) \ne \varphi$ occurs if and only if there exists a sequence of integers $i_1,\ldots,i_k$ so that $x_{i_1}\ldots x_{i_k} = y_{i_1}\ldots y_{i_k}$. Since the latter is the Post

Correspondence Problem and is known to be recursively unsolvable [8], whether $L(x,y)$ contains an u.p. sequence of period w is recursively unsolvable.

Remark. The family of languages $L(x)$ of Theorem 4.3 is also useful in showing the following questions about sequences and an arbitrary language L to be recursively unsolvable:

(1) Whether L contains a sequence containing a fixed letter. For let d be a letter not occurring in $L(x) + L(y)$ and consider $[\text{Init } L(x)](d^2)^* + [\text{Init } L(y)]d(d^2)^*$.

(2) Whether L contains a purely periodic sequence.(9) For let d be a letter not occurring in $L(x) + L(y)$ and consider $\text{Init } [dL(x)(dL(y))^*]$.

We conclude with a remark on the unsolvability of whether a language contains an u.p. sequence. Let D, L_1, and L_3 be as in Theorem 2.1. It follows from the proof of Theorem 2.1 that, for any word w in $\theta(a,b)$, $\text{Init } w \subseteq L_3$ if and only if w is in D. Let τ be the operation such that for each word u, $\tau(u) = u$ if the letter c occurs in u, and $\tau(u) = \varphi$ otherwise. If L is a language, then $\tau(L)$ is a language [5]. Thus $\tau(\text{Init } L(x))$ is a language. Consider the language

$$M(x,y) = L_3 + L(x)(d^2)^* + L(y)d(d^2)^* + \tau(\text{Init } L(x)).$$

Then $M(x,y)$ contains an u.p. sequence if and only if there exist integers $i_1,\ldots,i_k$, each $i_j \le n$, so that $x_{i_1}\cdots x_{i_k} = y_{i_1}\cdots y_{i_k}$ and this word is in D.

---

(9) A sequence of words $D = x_1\ldots x_n\ldots$ is said to be <u>purely periodic</u> if there exists an integer $m \ge 1$ so that $x_{i+m} = x_i$ for all $i \ge 1$.

In other words, whether a language contains an u.p. sequence is recursively unsolvable if the following modification of the Post Correspondence Problem is true: Given two n-tuples $(x_1,\ldots,x_n)$ and $(y_1,\ldots,y_n)$ of words in $\theta(a,b)-\epsilon$, it is recursively unsolvable whether there is a sequence of integers $i_1,\ldots,i_k$, each $i_j \leq n$, so that $x_{i_1}\ldots x_{i_k} = y_{i_1}\ldots y_{i_k}$ and $x_{i_1}\ldots x_{i_k}$ begins the sequence $bbaba^2ba^3\ldots$.

APPENDIX

Proof. We first recall some terminology and facts about generation trees. Let $G = (V,P,\Sigma,S)$ be a grammar. Call the elements of $V-\Sigma$ variables. Let w_1 be a variable. Let $w_2,\ldots,w_r$ be words in $\theta(V)$, $w_1 \to w_2$ a production, with the following property. For $2 \le i < r$ there exist words u_i, v_i, y_i, z_i such that $w_i = u_i y_i v_i$, $w_{i+1} = u_i z_i v_i$, and $y_i \to z_i$ is a production. A generation tree (constructed below) is a rooted, directed tree with an element of $V \cup \{\epsilon\}$, called the node name, associated at each node.

The nodes of the tree are certain tuples of the form $(i_1,\ldots,i_k)$, where $k \le r$ and i_j is a positive integer. The directed lines of the tree are all the ordered pairs $\langle(i_1,\ldots,i_k), (i_1,\ldots,i_k, i_{k+1})\rangle$ of nodes. Let the 1-tuple (1) be the root and w_1 the node name of (1). If $w_2 = \epsilon$ let (1,1) be a node in the tree and ϵ the node name of (1,1). If $w_2 = x_1^{(2)}\ldots x_{n(2)}^{(2)}$, each $x_i^{(2)}$ in V, let $(1,i)$, $1 \le i \le n(2)$, be a node and $x_i^{(2)}$ its node name. Continuing by induction, suppose that for all $t \le k$, every occurrence in w_t of an element of V serves as node name of some node. Now

$$(*) \quad u_k y_k v_k = w_k \Rightarrow w_{k+1} = u_k z_k v_k.$$

Let $(i_1,\ldots,i_s)$ be the node whose node name is the occurrence of y_k indicated in (*). If $z_k = \epsilon$ let $(i_1,\ldots,i_s,1)$ be a node and ϵ its node name. If $z_k = x_1^{(k)}\ldots x_{n(k)}^{(k)}$, each $x_i^{(k)}$ in V, let $(i_1,\ldots,i_s,i)$, $1 \le i \le n(k)$, be a node and $x_i^{(k)}$ its node name. This procedure is repeated through $k = r-1$. The resulting entity is the generation tree.

A node $(j_1,\ldots,j_t)$ is said to be an extension of the node $(i_1,\ldots,i_s)$ if $s \le t$ and $i_k = j_k$ for all $k \le s$.

A path in a generation tree is a sequence of nodes $N_1,\ldots,N_k$ such that $\langle N_i, N_{i+1}\rangle$ is a directed line for each $i \le k-1$.

Given the nodes $N_1 = (i_1,\ldots,i_s)$ and $N_2 = (j_1,\ldots,j_t)$ write $N_1 \le N_2$ if either N_2 is an extension of N_1 or if $i_k < j_k$ for the smallest integer k such that $i_k \ne j_k$.

The relation $\le$ is a simple order on the set of nodes.

A node is called maximal if there is no node distinct from it which is an extension of it.

We shall use (implicitly and explicitly) the following known facts about a generation tree T associated with $\xi \Rightarrow w$ [1]:

(a) If N is a nonmaximal node, then the node name x of N is a variable and $\xi \Rightarrow uxv$ for some u and v in $\theta(V)$.

(b) Let $N_1,\ldots,N_k$ be the maximal nodes, with $N_i \le N_{i+1}$ for each i. Then w is the word obtained by replacing in $N_1 \ldots N_k$ each node with its node name.

(c) Let N be a nonmaximal node in a generation tree, and x its node name. Then the "subtree" of T formed by using as nodes all extensions of N is a generation tree.

(d) Let $w = u\gamma v$ and let T_1 be a generation tree of $\gamma \Rightarrow w_1$. If T_1 is placed (in the obvious way) with its root on the node whose node name is γ in $u\gamma v$, then a generation tree of $\xi \Rightarrow uw_1v$ is obtained.

We now return to the proof of Theorem 2.5. Let D be a sequence satisfying the hypothesis of the theorem. Let L be any language containing D. We shall show that D contains an u.p. sequence. Consider the set $L' = L-\{\epsilon\}-\Sigma$. L' is a language and there is a grammar $G = (V,P,\Sigma,S)$, $L(G) = L'$, such that every production in P is of the form $\xi \to \mu\nu$, μ and ν in V [1]. Let N denote the number of distinct variables. Let H be the set of those variables ξ such that $\xi \Rightarrow a^s\xi a^t$ for some $s + t > 0$. Let H_1 be the set of those ξ in H such that $\xi \Rightarrow \xi a^t$ for some $t > 0$. [We can effectively determine H and H_1, but we do not need this fact.] We shall see below that H is nonempty. Denote the distinct elements of H by $\xi_1,\ldots,\xi_r$. For each ξ_i in H_1 let $e(i) > 0$ be an integer such that $\xi_i \Rightarrow \xi_i a^{e(i)}$. For each ξ_i in $H-H_1$ let $e(i) > 0$, $s(i)$, $t(i)$ be integers such that $e(i) = s(i) + t(i)$ and $\xi_i \Rightarrow a^{s(i)}\xi_i a^{t(i)}$. Let $e = e(1)\ldots e(r)$.

Consider any word ua^k in D, where $u \neq \epsilon$ and $k \geq 2^{(2N+e)|u|}$. We shall show that Init $(ua^*) \subseteq L$, thereby proving the theorem. Since Init $(ua^k) \subseteq D \subseteq L$, it suffices to show that ua^q is in L(G) for each $q > k$. Accordingly, let $q > k$ be given and let $p = |u|$. Then $k-2^{2Np} \geq 2^{(2N+e)p}-2^{2Np} = 2^{2Np}(2^{ep}-1) \geq 2(2^{ep}-1) \geq 2^{ep} > ep \geq e$. Therefore there is a positive integer g such that $2^{2Np} < q-ge \leq k$. Then ua^{q-ge} is in Init (ua^k) and $|ua^{q-ge}| \geq 2$. Thus ua^{q-ge} is in L(G). Hence there is a generation tree T of g which derives ua^{q-ge} (from S).

Since each production is of the form $\xi \to \mu\nu$, μ and ν in V, it is readily seen that any generation tree of G of a word of length $> 2^n$ contains a path with at least n+1 nodes, where each node name is a variable. Now $|ua^{q-ge}| > q-ge > 2^{2Np} \geq 2^{N(p+1)}$. Thus T contains a path $Z_1,\ldots,Z_{N(p+1)+1}$, where the node name of each

Z_i is a variable. Since there are only N distinct variables, one of them, say ξ, is the node name of at least p + 2 nodes. Denote by $Y_1,\ldots,Y_{p+2}$ the first p + 2 nodes in the path whose node name is ξ. For $1 \le i \le p+2$, let T_i be the subtree of T whose nodes are the extensions of Y_i. Then T_i is a generation tree (from ξ) of a word v_i in $\theta(\Sigma)-\epsilon$. For $1 \le i \le p+1$, since the node Y_{i+1} occurs in T_i, there are words x_i, y_i in $\theta(\Sigma)$ such that $\xi \Rightarrow x_i\xi y_i$ and $v_i = x_i v_{i+1} y_i$. Since each production is of the form $\gamma \to \mu\nu$, μ and ν in V, $x_i y_i \ne \epsilon$. Since Y_1 is in T, there exist w_1, w_2 in $\theta(V)$ such that $S \Rightarrow w_1\xi w_2$. Thus $ua^{q-ge} = w_1 x_1\cdots x_{p+1} v_{p+2} y_{p+1}\cdots y_1 w_2$.

Two cases arise.

(1) Suppose that one of the x_i is ϵ. Let j be the smallest integer such that $x_j = \epsilon$. Then $|x_1\ldots x_{j-1}| \ge j-1$. Since $x_i y_i \ne \epsilon$ for each i, $|x_{j+1}\cdots x_{p+1} v_{p+2} y_{p+1}\cdots y_{j+1}| \ge p+1-j$. Thus $|x_1\ldots y_{j+1}| \ge p$, so that u is an initial subword of $w_1x_1\ldots y_{j+1}$. Therefore y_j is in a^*. As $x_j = \epsilon$, $y_j \ne \epsilon$. Thus y_j is in aa^*. Since $\xi \Rightarrow x_j\xi y_j$, ξ is in H_1, say $\xi = \xi_d$. Now e is a multiple of e(d). Thus $\xi \Rightarrow \xi a^{e(d)ge/e(d)} = \xi a^{ge}$ and

$$\begin{aligned} S &\Rightarrow w_1 x_1\ldots x_j\xi y_j\ldots y_1 w_2 \\ &\Rightarrow w_1 x_1\ldots x_j \xi a^{ge} y_j\ldots y_1 w_2 \\ &\Rightarrow w_1 x_1\cdots x_{p+1} v_{p+2} y_{p+1}\cdots y_{j+1} a^{ge} y_j\ldots y_1 w_2 \\ &= ua^{q-ge}a^{ge} = ua^q. \end{aligned}$$

(2) Suppose that none of the x_i is ϵ. Then $|x_1\ldots x_p| \ge p$, so that u is an initial subword of $w_1x_1\ldots x_p$. Thus $x_{p+1}v_{p+2}y_{p+1}\cdots y_1 w_2$ is in aa^*. Then

$\xi \Rightarrow x_{p+1}\xi y_{p+1}$, with $x_{p+1}y_{p+1}$ in aa^*. Therefore ξ is in H, say $\xi = \xi_d$. Then there exist nonnegative integers s and t so that $\xi \Rightarrow a^s\xi a^t$ and $e(d) = s + t$. Thus

$$\begin{aligned} S &\Rightarrow v_1x_1\cdots x_p\xi y_p\cdots y_1v_2 \\ &\Rightarrow v_1x_1\cdots x_pa^{sge/e(d)}\xi a^{tge/e(d)}y_p\cdots y_1v_2 \\ &\Rightarrow v_1x_1\cdots x_pa^{sge/e(d)}x_{p+1}v_{p+2}y_{p+1}a^{tge/e(d)}y_p\cdots y_1v_2 \\ &= v_1x_1\cdots x_{p+1}v_{p+2}y_{p+1}\cdots y_1v_2a^{(s+t)ge/e(d)} \\ &\qquad \text{since } x_{p+1}\cdots y_1v_2a^{(s+t)ge/e(d)} \text{ is in } aa^* \\ &= ua^q. \end{aligned}$$

BIBLIOGRAPHY

[1] Bar-Hillel, Perles, and Shamir. "On Formal Properties of Simple Phrase Structure Grammars," Zeitschrift fur Phonetik, Sprachwissenschaft und Kommunikationsforschung, Vol. 14, 1961, pp.143-172.

[2] Chomsky, N. "Three Models for the Description of Language," IRE Transactions on Information Theory, Vol. 2, 1956, pp.113-124.

[3] Chomsky, N. "On Certain Formal Properties of Grammars," Information and Control, Vol. 2, 1959, pp.137-167.

[4] Ginsburg, S. and Rice, H. G. "Two Families of Languages Related to ALGOL," Journal of the Association for Computing Machinery, Vol. 9, 1962, pp.350-371.

[5] Ginsburg, S. and Rose, G. F. "Operations Which Preserve Definability in Languages," to appear in the Journal of the Association for Computing Machinery.

[6] Ginsburg, S. and Rose, G. F. "Some Recursively Unsolvable Problems in ALGOL-like Languages," Journal of the Association for Computing Machinery, Vol. 10, 1963, pp.29-47.

[7] Konig, D. Theorie der Endlichen und Unendlichen Graphen, Chelsea Publishing Company, New York, 1950.

[8] Post, E. L. "A Variant of a Recursively Unsolvable Problem," Bulletin of the American Mathematical Society, Vol. 52, 1946, pp.264-268.

[9] Rabin, M. and Scott, D. "Finite Automata and Their Decision Problems," IBM Journal of Research and Development, Vol. 3, 1959, pp.114-125.

Distribution List

Code	Organization	No. of Copies
AF 5	AFMTC (AFMTC Tech Library-MU-135) Patrick AFB, Fla.	1
AF 18	AUL Maxwell AFB, Ala.	1
AF 32	OAR (RROS, Col. John R. Fowler) Tempo D 4th and Independence Ave, Wash 25, D. C.	1
AF 33	AFOSR, OAR (SRYP) Tempo D 4th and Independence Ave, Wash 25, D. C.	1
AF 43	ASD (ASAPRD - Dist) Wright-Patterson AFB, Ohio	1
AF 124	RADC (RAAID) Griffiss AFB, New York Attn: Documents Library	1
AF 139	AF Missile Development Center (MDGRT) Holloman AFB, New Mexico	1
AF 314	Hq. OAR (RROSP, Maj. Richard W. Nelson) Washington 25, D. C.	1
AF 318	ARL (ARA-2) Library AFL 2292, Building 450 Wright-Patterson AFB, Ohio	1
Ar 5	Commanding General USASRDL Ft. Monmouth, N.J. Attn: Tech Doc. Ctr. SIGRA/SL-ADT	1
Ar 9	Department of the Army Office of the Chief Signal Officer Washington 25, D. C. Attn: SIGRD-4a-2	1

Code	Organization	No. of Copies
Ar 50	Commanding Officer Attn: ORDTL-012 Diamond Ordnance Fuze Laboratories Washington 25, D. C.	1
Ar 67	Redstone Scientific Information Center U.S. Army Missile Command Redstone Arsenal, Alabama	1
G 31	Office of Scientific Intelligence Central Intelligence Agency 2430 E Street, N.W. Washington 25, D. C.	1
G 2	ASTIA (TIPAA) Arlington Hall Station Arlington 12, Virginia	20
G 68	Scientific and Technical Information Facility Attn: NASA Representative (S-AK/DL) P.O. Box 5700 Bethesda, Maryland	1
G 109	Director Langley Research Center National Aeronautics and Space Administration Langley Field, Virginia	
N 9	Chief, Bureau of Naval Weapons Department of the Navy Washington 25, D. C. Attn: DLI-31	2
N 29	Director (Code 2027) U.S. Naval Research Laboratory Washington 25, D. C.	2
I 292	Director, USAF Project RAND The RAND Corporation 1700 Main Street Santa Monica, California THRU: AF Liaison Office	1

Code	Organization	No. of Copies
M 6	AFCRL, OAR (CRXRA - Stop 39) L. G. Hanscom Field Bedford, Mass.	20
AF 253	Technical Information Office European Office, Aerospace Research Shell Building, 47 Cantersteen Brussels, Belgium	1
Ar 107	U.S. Army Aviation Human Research Unit U.S. Continental Army Command P.O. Box 428, Fort Rucker, Alabama Attn: Maj. Arne H. Eliasson	1
G 8	Library Boulder Laboratories National Bureau of Standards Boulder, Colorado	2
M 63	Institute of the Aerospace Sciences, Inc. 2 East 64th Street New York 21, New York Attn: Librarian	1
M 84	AFCRL, OAR (CRXR, J. R. Marple) L. G. Hanscom Field Bedford, Mass.	1
N 73	Office of Naval Research Branch Office, London Navy 100, Box 39 F.P.O., New York, N.Y.	10
U 32	Massachusetts Institute of Technology Research Laboratory Building 26, Room 327 Cambridge 39, Massachusetts Attn: John H. Hewitt	1
U 431	Alderman Library University of Virginia Charlottesville, Virginia	1

Code	Organization	No. of Copies
G 9	Defence Research Member Canadian Joint Staff 2450 Massachusetts Avenue, N.W. Washington 8, D. C.	1
	Hq. AFCRL, OAR (CRRBA, A. Lewis Bastian, Jr.) L. G. Hanscom Field Bedford, Massachusetts	20
	System Development Corporation	
	S. B. Ginsburg (Room 9531)	100
	L. F. McConnell (Room 10156)	1
	C. Baum (Room 9012A)	8

Air Force Cambridge Research Laboratories, Bedford, Massachusetts.
AFCRL SOME REMARKS ABOUT SEQUENCES IN CONTEXT FREE LANGUAGES.
Scientific report no. 2,
31 January 1963, 31p., 9 refs.
Unclassified report

The following conjecture is considered:
(*) It is unsolvable whether a language (= context free language) contains a sequence.
While this conjecture is left unresolved, a number of results

1. Language.
2. Algebras.
3. Automata.
I. Project no. 4641, Task no. 46410.
II. Contract AF 19(628)-485.
III. System Development Corporation, Santa Monica, California.
IV. Ginsburg, S., Ullian, J. S.

Air Force Cambridge Research Laboratories, Bedford, Massachusetts.
AFCRL SOME REMARKS ABOUT SEQUENCES IN CONTEXT FREE LANGUAGES.
Scientific report no. 2,
31 January 1963, 31p., 9 refs.
Unclassified report

The following conjecture is considered:
(*) It is unsolvable whether a language (= context free language) contains a sequence.
While this conjecture is left unresolved, a number of results

1. Language.
2. Algebras.
3. Automata.
I. Project no. 4641, Task no. 46410.
II. Contract AF 19(628)-485.
III. System Development Corporation, Santa Monica, California.
IV. Ginsburg, S., Ullian, J. S.

Air Force Cambridge Research Laboratories, Bedford, Massachusetts.
AFCRL SOME REMARKS ABOUT SEQUENCES IN CONTEXT FREE LANGUAGES.
Scientific report no. 2,
31 January 1963, 31p., 9 refs.
Unclassified report

The following conjecture is considered:
(*) It is unsolvable where a language (= context free language) contains a sequence.
While this conjecture is left unresolved, a number of results

1. Language.
2. Algebras.
3. Automata.
I. Project no. 4641, Task no. 46410.
II. Contract AF 19(628)-485.
III. System Development Corporation, Santa Monica, California.
IV. Ginsburg, S., Ullian, J. S.

Air Force Cambridge Research Laboratories, Bedford, Massachusetts.
AFCRL SOME REMARKS ABOUT SEQUENCES IN CONTEXT FREE LANGUAGES.
Scientific report no. 2,
31 January 1963, 31p., 9 refs.
Unclassified report

The following conjecture is considered:
(*) It is unsolvable whether a language (= context free language) contains a sequence.
While this conjecture is left unresolved, a number of results

1. Language.
2. Algebras.
3. Automata.
I. Project no. 4641, Task no. 46410.
II. Contract AF 19(628)-485.
III. System Development Corporation, Santa Monica, California.
IV. Ginsburg, S., Ullian, J. S.

pertaining to it are obtained. For example, the unsolvability of whether a language contains the set $\{aba^2ba^3 \ldots ba^n / n \geq 1\}$, implies (*). It is shown that (*) is equivalent to the unsolvability of whether a language contains a chain of a special form. Several facts about whether a language contains a specific sequence are also demonstrated. In particular, it is shown that whether a language contains a given sequence is unsolvable, but whether a language contains a given ultimately periodic sequence is solvable.

V. TM-738/001/00.
VI. In ASTIA Collection.

pertaining to it are obtained. For example, the unsolvability of whether a language contains the set $\{aba^2ba^3 \ldots ba^n / n \geq 1\}$, implies (*). It is shown that (*) is equivalent to the unsolvability of whether a language contains a chain of a special form. Several facts about whether a language contains a specific sequence are also demonstrated. In particular, it is shown that whether a language contains a given sequence is unsolvable, but whether a language contains a given ultimately periodic sequence is solvable.

V. TM-738/001/00.
VI. In ASTIA Collection.

pertaining to it are obtained. For example, the unsolvability of whether a language contains the set $\{aba^2ba^3 \ldots ba^n / n \geq 1\}$, implies (*). It is shown that (*) is equivalent to the unsolvability of whether a language contains a chain of a special form. Several facts about whether a language contains a specific sequence are also demonstrated. In particular, it is shown that whether a language contains a given sequence is unsolvable, but whether a language contains a given ultimately periodic sequence is solvable.

V. TM-738/001/00.
VI. In ASTIA Collection.

pertaining to it are obtained. For example, the unsolvability of whether a language contains the set $\{aba^2ba^3 \ldots ba^n / n \geq 1\}$, implies (*). It is shown that (*) is equivalent to the unsolvability of whether a language contains a chain of a special form. Several facts about whether a language contains a specific sequence are also demonstrated. In particular, it is shown that whether a language contains a given sequence is unsolvable, but whether a language contains a given ultimately periodic sequence is solvable.

V. TM-738/001/00.
VI. In ASTIA Collection.